

Optimal Structured Light à la Carte: Supplemental Document

Parsa Mirdehghan Wenzheng Chen Kiriakos N. Kutulakos
Department of Computer Science
University of Toronto
{parsa, wenzheng, kyros}@cs.toronto.edu

A. ZNCC Optimality Proof

Proposition 1 (ZNCC Decoding in presence of ambient light). If observation vectors and code vectors are related according to:

$$\mathbf{o}_q = \mathbf{T}[p, q] \cdot \mathbf{c}_p + a_q + \mathbf{e}_q \quad (1)$$

Then

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \arg \max_{1 \leq p \leq N} \text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p) \quad (2)$$

where $\text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p)$ is the zero-mean normalized cross-correlation of two vectors, and v is the variance of the variances of the code vectors:

$$\text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p) = \frac{\mathbf{o}_q - \text{mean}(\mathbf{o}_q)}{\|\mathbf{o}_q - \text{mean}(\mathbf{o}_q)\|} \cdot \frac{\mathbf{c}_p - \text{mean}(\mathbf{c}_p)}{\|\mathbf{c}_p - \text{mean}(\mathbf{c}_p)\|}, \quad v = \text{var}(\{\text{var}(\mathbf{c}_1), \dots, \text{var}(\mathbf{c}_N)\}) \quad (3)$$

Proof of Proposition 1.

Let us define L_q^p as the likelihood of observing \mathbf{o}_q given that pixel q 's true correspondence is projector pixel p :

$$\begin{aligned} L_q^p &= \Pr(\mathbf{o}_q | \mathbf{c}_p) \\ &= \int_0^X \int_0^Y \Pr(\mathbf{o}_q | \mathbf{c}_p, \mathbf{T}[p, q]=x, a_q=y) \Pr(x) \Pr(y) dy dx \end{aligned} \quad (4)$$

Based on our assumption, discussed in Section 3 of the paper, $\mathbf{T}[p, q]$ and ambient probabilities, $\Pr(x)$ and $\Pr(y)$, correspond to uniform distributions over $[0, X]$ and $[0, Y]$ respectively. Following the image formation model in Eq. (1), we can write the likelihood as following:

$$L_q^p = \int_0^X \int_0^Y \Pr(\mathbf{e}_q = \mathbf{o}_q - x\mathbf{c}_p - y\mathbf{1}) \frac{1}{XY} dy dx \quad (5)$$

We assumed the noise vector \mathbf{e}_q contains i.i.d Gaussian distributions with standard deviation σ (refer to Section 3 of the main paper); therefore the likelihood can be further expanded as:

$$L_q^p = \frac{1}{XY} \int_0^X \int_0^Y (\sqrt{2\pi}\sigma)^{-K} \exp\left(\left(\frac{-1}{2\sigma^2}\right)(\mathbf{o}_q - x\mathbf{c}_p - y\mathbf{1})^t(\mathbf{o}_q - x\mathbf{c}_p - y\mathbf{1})\right) dy dx \quad (6)$$

Where $()^t$ denotes the transpose operator. By denoting the constant terms with G , and expanding the expression inside the exponential function, we have:

$$L_q^p = G \int_0^X \int_0^Y \exp\left(\left(\frac{-1}{2\sigma^2}\right)(\mathbf{o}_q^t \mathbf{o}_q - 2x\mathbf{o}_q^t \mathbf{c}_p - 2y\mathbf{o}_q^t \mathbf{1} + x^2 \mathbf{c}_p^t \mathbf{c}_p + 2xy \mathbf{c}_p^t \mathbf{1} + Ky^2)\right) dy dx \quad (7)$$

Before continuing the proof of Proposition 1, we first consider the special case of image formation model expressed in Eq. (1), where the ambient contribution for every camera pixel is zero.

Lemma 1 (NCC Decoding in absence of ambient light). If observation vectors and code vectors are related according to:

$$\mathbf{o}_q = \mathbf{T}[p, q] \cdot \mathbf{c}_p + \mathbf{e}_q \quad (8)$$

Then

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \arg \max_{1 \leq p \leq N} \text{NCC}(\mathbf{o}_q, \mathbf{c}_p) \quad (9)$$

where $\text{NCC}(\mathbf{o}_q, \mathbf{c}_p)$ denotes the normalized cross-correlation of two vectors and v is the variance of code vectors' norms:

$$\text{NCC}(\mathbf{o}_q, \mathbf{c}_p) = \frac{\mathbf{o}_q}{\|\mathbf{o}_q\|} \cdot \frac{\mathbf{c}_p}{\|\mathbf{c}_p\|} \quad , \quad v = \text{var}(\{\|\mathbf{c}_1\|, \dots, \|\mathbf{c}_N\|\}) \quad (10)$$

Proof of Lemma 1. The general expression for the likelihood described in Eq. (7) can be simplified into:

$$L_q^p = G \int_0^X \exp \left(\left(\frac{-1}{2\sigma^2} \right) (\mathbf{o}_q^t \mathbf{o}_q + x^2 \mathbf{c}_p^t \mathbf{c}_p - 2x \mathbf{o}_q^t \mathbf{c}_p) \right) dx \quad (11)$$

By taking the constant term $\mathbf{o}_q^t \mathbf{o}_q$ out of the integral and completing the square inside the exponential function, we have:

$$L_q^p = G \exp \left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2} \right) \exp \left(\frac{(\mathbf{o}_q^t \mathbf{c}_p)^2}{2\sigma^2 \|\mathbf{c}_p\|^2} \right) \int_0^X \exp \left(- \left(\frac{x \|\mathbf{c}_p\|}{\sqrt{2}\sigma^2} - \frac{\mathbf{o}_q^t \mathbf{c}_p}{\sqrt{2}\sigma^2 \|\mathbf{c}_p\|} \right)^2 \right) dx \quad (12)$$

The integral is related to the *error function*: $\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x \exp(-t^2) dt$, and does not have closed-form expression. Using this function, the integral in Eq. (12) can be written as:

$$L_q^p = \left(\frac{1}{X} \right) (\sqrt{2\pi}\sigma)^{-K} \exp \left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2} \right) \underbrace{\exp \left(\frac{(\mathbf{o}_q^t \mathbf{c}_p)^2}{2\sigma^2 \|\mathbf{c}_p\|^2} \right)}_{A(\sigma)} \underbrace{\left(\frac{\sqrt{\pi}\sigma}{\sqrt{2}\|\mathbf{c}_p\|} \right)}_{B(\sigma)} \left[\underbrace{\text{erf} \left(\frac{X \|\mathbf{c}_p\| - \mathbf{o}_q^t \mathbf{c}_p}{\sqrt{2}\sigma \|\mathbf{c}_p\|} \right)}_{C(\sigma)} - \underbrace{\text{erf} \left(\frac{-\mathbf{o}_q^t \mathbf{c}_p}{\sqrt{2}\sigma \|\mathbf{c}_p\|} \right)}_{D(\sigma)} \right] \quad (13)$$

The error function $\text{erf}(x)$ goes to 1 for large positive x , and goes to -1 for large negative x (Figure 1). Therefore, when the noise standard deviation σ , goes to zero, we have $\lim_{\sigma \rightarrow 0} C(\sigma) = 1$ and $\lim_{\sigma \rightarrow 0} D(\sigma) = -1$, and then we can conclude $\lim_{\sigma \rightarrow 0} C(\sigma) - D(\sigma) = 2$. Furthermore, as the lengths of different code vectors get closer to each other, or strictly speaking $\text{var}(\{\|\mathbf{c}_1\|, \dots, \|\mathbf{c}_N\|\}) \rightarrow 0$, term $B(\sigma)$ will be the same in the likelihood of different projector pixels. Therefore, the likelihood in Eq. (13) will be dominated by term $A(\sigma)$ and the ML solution is given by:

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \arg \max_{1 \leq p \leq N} \exp \left(\frac{(\mathbf{o}_q^t \mathbf{c}_p)^2}{2\sigma^2 \|\mathbf{c}_p\|^2} \right) = \arg \max_{1 \leq p \leq N} \text{NCC}(\mathbf{o}_q, \mathbf{c}_p) \quad (14)$$

QED. \square

Proof of Proposition 1 (Continued). By rearranging the general expression for likelihood in Eq. (7), we have:

$$L_q^p = G \exp \left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2} \right) \int_0^X \exp \left(\left(\frac{-1}{2\sigma^2} \right) (-2x \mathbf{o}_q^t \mathbf{c}_p + x^2 \mathbf{c}_p^t \mathbf{c}_p) \right) \underbrace{\left[\int_0^Y \exp \left(\left(\frac{-1}{2\sigma^2} \right) (-2y \mathbf{o}_q^t \mathbf{1} + 2xy \mathbf{c}_p^t \mathbf{1} + Ky^2) \right) dy \right]}_H dx \quad (15)$$

Let us denote the inner integral inside the bracket by H . As in the proof of Lemma 1, we express H by completing the square

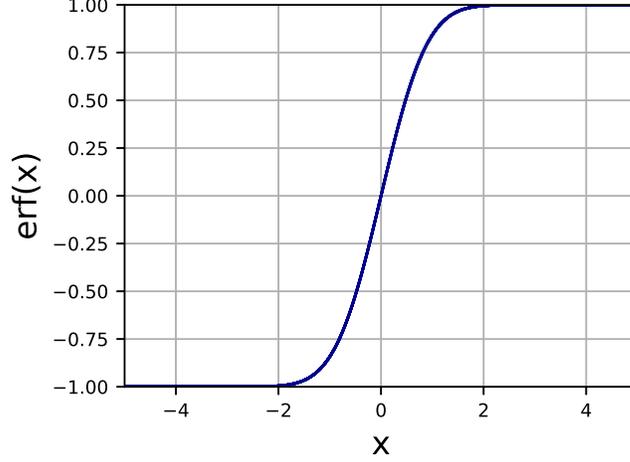


Figure 1: **Error function.** $\text{erf}(x)$ becomes 1 and -1 for large positive and large negative values respectively.

inside the exponential function, and simplify it as two error functions:

$$\begin{aligned}
 H &= \exp\left(\frac{(\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1})^2}{2K\sigma^2}\right) \int_0^Y \exp\left(-\left(\frac{\sqrt{K}y}{\sqrt{2}\sigma} - \frac{\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1}}{\sqrt{2K}\sigma}\right)^2\right) dy \\
 &= \exp\left(\frac{(\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1})^2}{2K\sigma^2}\right) \frac{\sqrt{2}\sigma}{\sqrt{K}} \frac{\sqrt{\pi}}{2} \left[\text{erf}\left(\frac{KY - (\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1})}{\sqrt{2K}\sigma}\right) - \text{erf}\left(-\frac{(\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1})}{\sqrt{2K}\sigma}\right) \right]
 \end{aligned} \tag{16}$$

Using the same argument discussed in Lemma 1 for simplifying Eq. (13), as σ goes to zero the difference between error functions goes to 2. Therefore, we can substitute term H in Eq. (15) to compute the integral over x :

$$\begin{aligned}
 L_q^p \underset{\sigma \rightarrow 0}{=} & G \exp\left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2}\right) \frac{\sqrt{2\pi}\sigma}{\sqrt{K}} \int_0^X \exp\left(\left(\frac{-1}{2\sigma^2}\right)\left(-2x\mathbf{o}_q^t \mathbf{c}_p + x^2 \mathbf{c}_p^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1} - x \mathbf{c}_p^t \mathbf{1})^2}{K}\right)\right) dx \\
 \underset{\sigma \rightarrow 0}{=} & G \exp\left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2}\right) \frac{\sqrt{2\pi}\sigma}{\sqrt{K}} \exp\left(\frac{(\mathbf{o}_q^t \mathbf{1})^2}{2K\sigma^2}\right) \exp\left(\frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})^2}{2\sigma^2(\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K})}\right) \\
 & \underbrace{\int_0^X \exp\left(-\left(\frac{x\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}{\sqrt{2}\sigma} - \frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})}{\sqrt{2}\sigma\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}\right)^2\right) dx}_I
 \end{aligned} \tag{17}$$

By completing the square inside the exponential function in I , and expressing it as two error functions, we have:

$$I = \left(\frac{\sqrt{\pi}}{2}\right) \left(\frac{\sqrt{2}\sigma}{\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}\right) \left[\text{erf}\left(\frac{X(\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}) - (\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})}{\sqrt{2}\sigma\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}\right) - \text{erf}\left(-\frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})}{\sqrt{2}\sigma\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}\right) \right] \tag{18}$$

As noise standard deviation σ goes to zero, the difference between error functions inside I goes to 2. Therefore, the likelihood

function will be:

$$L_q^p \underset{\sigma \rightarrow 0}{=} \left(\frac{1}{XY}\right) (\sqrt{2\pi}\sigma)^{-K} \left(\frac{\sqrt{2}\pi\sigma}{\sqrt{K}}\right) \underbrace{\left(\frac{\sqrt{2}\sigma}{\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}}\right)}_{U(\sigma)} \exp\left(\frac{-\|\mathbf{o}_q\|^2}{2\sigma^2}\right) \exp\left(\frac{(\mathbf{o}_q^t \mathbf{1})^2}{2K\sigma^2}\right) \underbrace{\exp\left(\frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})^2}{2\sigma^2(\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K})}\right)}_{V(\sigma)} \quad (19)$$

To better understand Eq. (19), we will simplify two terms:

$$\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K} = \left(\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1}\right)^t \left(\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1}\right) = K \text{var}(\mathbf{c}_p) \quad (20)$$

$$\frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})}{\sqrt{\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K}}} = \frac{(\mathbf{o}_q - \frac{(\mathbf{o}_q^t \mathbf{1})}{K} \mathbf{1})^t (\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1})}{\sqrt{(\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1})^t (\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1})}} = \|\mathbf{o}_q - \text{mean}(\mathbf{o}_q)\| \cdot \text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p) \quad (21)$$

It is worth mentioning that $\frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1}$ represents a constant vector containing the mean value of \mathbf{c}_p , and $(\mathbf{c}_p - \frac{(\mathbf{c}_p^t \mathbf{1})}{K} \mathbf{1})$ is the average-subtracted code vector at pixel p .

Based on Eq. (20), as the variances of code vectors get closer to each other, or in other word $\text{var}(\{\text{var}(\mathbf{c}_1), \dots, \text{var}(\mathbf{c}_N)\}) \rightarrow 0$, term $U(\sigma)$ in the likelihood of different correspondences will be the same, and the projector pixel that maximizes term $V(\sigma)$, has maximum likelihood; and also based on Eq. (21), term $V(\sigma)$ is a monotonic function of $\text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p)$. Thus, we can conclude:

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \arg \max_{1 \leq p \leq N} \exp\left(\frac{(\mathbf{o}_q^t \mathbf{c}_p - \frac{(\mathbf{o}_q^t \mathbf{1})(\mathbf{c}_p^t \mathbf{1})}{K})^2}{2\sigma^2(\|\mathbf{c}_p\|^2 - \frac{(\mathbf{c}_p^t \mathbf{1})^2}{K})}\right) = \arg \max_{1 \leq p \leq N} \text{ZNCC}(\mathbf{o}_q, \mathbf{c}_p) \quad (22)$$

QED. \square

B. Incorporating Projector Defocus

B.1. The Defocus Blur Coefficients

We assume that the only source of blur is projector defocus, *i.e.*, the camera's aperture is small enough to ensure that scene points are imaged blur-free on the camera's image plane. In this case, the 3D scene point corresponding to projector pixel p and camera pixel q will be illuminated by a defocused projection pattern. In the epipolar-only image formation model this defocus occurs exclusively on the epipolar plane, *i.e.*, blur can be expressed as a 1D blur kernel along the epipolar line.

Consider a projector lens with aperture A and focal length F that follows the thin-lens model. The blur radius at projector pixel p is given by (Figure 2):

$$\rho = A \frac{|f_2 - f_1|}{f_1} \quad (23)$$

where f_2 is the distance of the aperture from the projector's image plane and f_1 is the aperture's distance from the plane that brings into focus the scene point projecting to pixel p . According to the thin-lens model,

$$f_1 = \frac{d_1 F}{d_1 - F} \quad (24)$$

where d_1 is the depth of the scene point projecting to projector pixel p and camera pixel q .

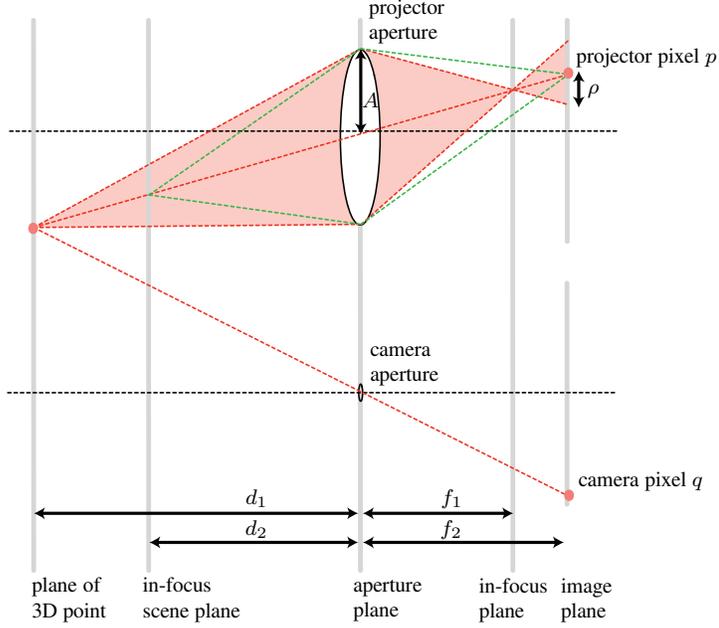


Figure 2: Thin-lens model of projector defocus.

We model the blur kernel as a 1D pillbox function whose radius is the radius of the blur circle:

$$\mathcal{K}(x) = \begin{cases} \frac{1}{2\rho} & \text{if } x \leq \rho \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

Assuming for simplicity that ρ is measured in units of projector pixels and that it takes on discrete values, we obtain the following expression for the coefficients b_i^{pq} in Eq. (7) of the paper:

$$b_i^{pq} = \begin{cases} \frac{1}{2\rho} & \text{if } |i - p| \leq \rho \\ 0 & \text{otherwise.} \end{cases} \quad (26)$$

B.2. Optimality of Defocused ZNCC Decoding.

Corollary 1 (Defocused ZNCC Decoder). If observation vectors and code vectors are related according to Eq. (7) in the paper, then:

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \text{Decode}(\mathbf{o}_q, \mathbf{C}\mathbf{T}^q) \quad (27)$$

Proof. Using the same notation as ZNCC optimality proof, and following the image formation model in Eq. (7) of the paper, likelihood L_q^p can be written as:

$$L_q^p = \int_0^X \int_0^Y \Pr(\mathbf{e}_q = \mathbf{o}_q - x \left(\sum_{i=1}^N b_i^{pq} \mathbf{c}_i \right) - y\mathbf{1}) \frac{1}{XY} dy dx \quad (28)$$

By denoting $(\sum_{i=1}^N b_i^{pq} \mathbf{c}_i)$ with $\bar{\mathbf{c}}_{pq}$, we can write:

$$L_q^p = \int_0^X \int_0^Y \Pr(\mathbf{e}_q = \mathbf{o}_q - x\bar{\mathbf{c}}_{pq} - y\mathbf{1}) \frac{1}{XY} dy dx \quad (29)$$

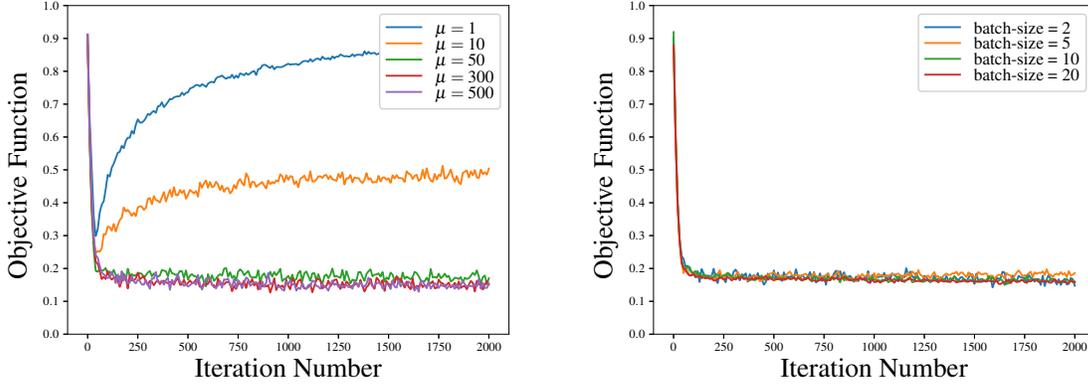


Figure 3: **Hyper-parameter tuning** Validation error (on 500 fixed random samples) over iterations in optimization of a sample code matrix of 4 patterns and 608 pixels with maximum frequency 16.

which has the same form as Eq. (5) in this supplemental document. Following exactly the same steps as proof for Proposition 1, we will have:

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \arg \max_{1 \leq p \leq N} \text{ZNCC}(\mathbf{o}_q, \bar{\mathbf{c}}_{pq}) \quad (30)$$

Based on the definition of \mathbf{T}^q in section 4 of the paper, $\bar{\mathbf{c}}_{pq}$ is the p 's column of \mathbf{CT}^q . Therefore, based on the ZNCC decoding definition in Eq. (11) of the paper, we can conclude:

$$\lim_{\substack{v \rightarrow 0 \\ \sigma \rightarrow 0}} \left(\arg \max_{1 \leq p \leq N} \Pr(\mathbf{o}_q | \mathbf{c}_p) \right) = \text{Decode}(\mathbf{o}_q, \mathbf{CT}^q) \quad (31)$$

QED. \square

C. Optimization Hyper-parameters

Two hyper-parameters require tuning in our framework: (1) the multiplier μ in the softmax approximation of Eq. (17) in the main paper, and (2) the mini-batch size for performing stochastic gradient descent. Figure 3 shows the effect of these two parameters in optimizing a code matrix with 4 patterns and maximum frequency 16, without any geometry constraints.

As can be seen from Figure 3 (left), when the multiplier μ decreases, the softmax approximation of the objective function is less accurate. As a result, the optimization does not lead to a good local minimum. On the other hand, increasing the multiplier beyond 300 does not have any noticeable impact on minimizing the objective function. Thus, to avoid any floating-point arithmetic issues and to have more stable gradients, we picked $\mu = 300$ for all of our code optimizations.

To perform the gradient descent, in each iteration we draw random samples of \mathbf{T} , \mathbf{E} , and \mathbf{a} , where each sample includes all the valid camera pixels. It can be seen from Figure 3 (right) that increasing the sample size does not affect the attained minimum of the objective function over the pre-drawn validation set. Therefore, to speed up the optimization process we used 2-sample mini-batches for all our code optimizations.

To gain some intuition about why the mini-batch size does not affect the results, consider the following. Since in each iteration we draw new samples of \mathbf{T} , \mathbf{E} , and \mathbf{a} , we are essentially training on an infinite-sized dataset, which helps avoid overfitting. As a result, even small mini-batch sizes can still converge to good local minima.

Together, the fast convergence rate of our optimization—less than 250 iterations—and the ability to use very small mini-batch sizes for gradient computation, allow us to optimize code matrices at near-interactive rates.

D. Additional Experimental Results

All the results discussed below were obtained with the experimental system discussed in Section 6 of the main paper. The setup is shown in Figure 7 (row 2, col 3) of the main paper. We performed three sets of experiments:

- Reconstructing the object shown in Figure 8 top row (“statue”) with many different optimized code matrices, and comparing the depth maps to each other and to those obtained with MPS [1] and EPS [2] codes matrices. Here we report results from a total of 58 individual 3D acquisition experiments (refer to the project web page [3] for an identical suite of experiments with the object shown in Figure 1).
- Reconstructing cross-sections of an additional object that exhibit significant subsurface scattering. We used epipolar-only imaging for this set of experiments, and compare the results from our optimized code matrices to those obtained by MPS and EPS code matrices under epipolar-only imaging conditions. Here we report results from another 59 individual 3D acquisition experiments (see the project web page [3] for a similar set of experiments with an object exhibiting inter-reflection).
- We also show additional quantitative comparisons that extend the results in Figure 7 of the main paper by taking projector defocus into account.

D.1. Experimental Evaluation of Optimal Code Matrices

Figures 6-10 show reconstruction results for a large set of 3D acquisition experiments. Each experiment uses a different code matrix, *i.e.*, a different sequence of projection patterns. A single experiment consists of projecting a sequence of patterns one by one onto the scene, capturing the resulting images, and using ZNCC decoding to establish correspondences between camera pixels and projector pixels. In each case, we compare the results to those obtained by MPS and EPS codes, respectively, and to a pattern sequence with more than $20\times$ the number of patterns. The set of code matrices shown in Figures 6-10 corresponds to a “walk” in the space of optimal codes, as in Figure 6 of the main paper.

Our choice of ZNCC decoding for all depth map calculations is justified on two grounds: (1) the goal of these experiments is to assess the quality of the optimized codes themselves, not the decoding algorithm and (2) as shown in Figure 5 of the main paper, ZNCC decoding yields comparable results to native decoders.

Acquiring “ground-truth” disparities To evaluate the reconstructed disparities both qualitatively and quantitatively, we compare them to disparities computed by a completely different and much longer pattern sequence: 160 phase-shifted sinusoidal projection patterns (10 shifts of 16 sinusoidal patterns, with frequencies 1 through 16). We decode the 160 captured images by ZNCC decoding to assign a depth to each pixel. We manually segmented this ground-truth depth map to create a binary mask that indicates which pixels actually lie on the object. All these pixels were marked as having a valid ground-truth depth; the remaining pixels correspond to pixels that were on the black backdrop onto which the object was placed, and were marked as having an unknown ground-truth depth. The resulting binary mask is shown in the top row of Figure 6.

Visualizing the results of a single 3D acquisition experiment Figure 5 describes how we present the results of an individual acquisition experiment. To qualitatively assess the performance of a code matrix, we show the *raw* depth maps computed when using it. To assess it quantitatively, we compare 2D slices of the reconstructed 3D pointset against those computed by the ground-truth sequence, and show the histogram of differences in the computed and ground-truth disparities.

Code performance as a function of maximum spatial frequency Figure 6 shows 3D acquisition results for a different portion of the wooden object shown in Figure 8 of the main paper. Across all encoding schemes, the choice of maximum frequency has a big impact on the reconstructed depth maps, both in the 4-pattern and 5-pattern experiments. These results suggest that our optimized codes produce better reconstructions overall, with a bigger fraction of the pixel disparities having a ground-truth error of a pixel or less, and fewer outliers overall. It is also important to note that our codes are able to exploit higher frequencies more effectively, with the best performance observed at higher spatial frequencies than MPS and EPS codes. This is significant given that both MPS and EPS codes were specifically designed to enable 3D reconstruction using pattern sequences with high spatial frequencies.

Code performance as a function of camera noise model Figure 7 shows 3D acquisition results using codes that were optimized for three different noise models. Of the three, the exponential-plus-additive-Gaussian model is the most pessimistic about the effect of signal-dependent noise: unlike Poisson shot noise whose variance increases linearly with the

signal, noise variance in the exponential model increases quadratically with the signal. Nevertheless, despite the very significant difference between the three noise models, reconstruction results for all three noise-optimized codes were very similar. While the Poisson-plus-additive-Gaussian codes yielded error histograms which were marginally better for four patterns—as would be expected from a model that captures the actual noise statistics of a conventional CMOS sensor under incoherent illumination—the effect is relatively small. We decided to fix the model to exponential-plus-additive-Gaussian for all subsequent code optimizations. This intentionally gave a slight disadvantage to our optimized codes.

Code performance as a function of error tolerance Figure 8 shows 3D acquisition results using codes that were optimized for four different tolerances in disparity error. Perhaps counterintuitively, codes optimized for one- and two-pixel error tolerance performed considerably better than codes optimized for perfect correspondences: almost 50% of the pixels were within a pixel of the true disparity for optimized 4-pattern codes, and almost 80% of them for optimized 5-pattern codes. This also significantly increased the performance margin over the MPS and EPS results shown in Figure 6.

Code performance as a function of geometric constraints Figure 9 shows 3D acquisition results using codes that were optimized for three different geometric constraints. Unlike the experiments in Figures 6-8, *a priori* geometric information can improve the decoding of any coding scheme, not just our optimized codes. In particular, during decoding we only search the set of valid correspondences, as specified by the geometry matrix \mathbf{G} , and choose the one whose code maximizes its ZNCC score. Compared with the baseline reconstruction results in Figure 6, this improves the depth map across all coding schemes, frequencies and number of patterns. The reconstruction results from our geometry-optimized codes, however, is significantly better. This indicates that applying these constraints both for code optimization and for decoding yields superior performance.

Code optimization for 3-pattern sequences In theory, our optimization scheme can be used to generate optimized code matrices for just three projection patterns. As shown in Figure 10, however, we found these matrices to be less effective than conventional 3-pattern phase shifting. While the reason behind this sub-optimal performance is unclear, we believe that the available degrees of freedom in code matrix design are probably too small to yield an improvement over conventional phase shifting.

D.2. Experiments under Epipolar-Only Imaging Conditions

To evaluate our codes’ performance for objects that exhibit significant indirect light transport we performed preliminary epipolar-only imaging experiment with a translucent candle. This object is most appropriately imaged with an epipolar-only imaging system [4, 5]. These systems block all light transport paths that lie outside the epipolar plane in order to diminish their influence on the reconstructed geometry. In lieu of such a system, we show reconstruction results for just 3 epipolar slices of this object. Each slice was reconstructed by simulating the epipolar-only imaging procedure as follows:

1. *The i -th projection pattern:* we define a 2D pattern that is zero everywhere except along one specific epipolar line on the projector plane. The pattern along that epipolar line is given by the i -th row of the code matrix.¹
2. *The i -th image:* after capturing an image of the scene under the i -th projection pattern, we extract the pixels along the corresponding epipolar line on the camera plane.² These pixels are the epipolar-only image of the scene’s intersection with the epipolar plane defined by the chosen epipolar line.
3. *Reconstruction procedure:* After acquiring K images of that intersection, we perform ZNCC decoding as in the previous experiments.

For each code matrix, we repeated this procedure three times to reconstruct three different epipolar slices of the object. As in the experiments of the previous section, we obtained ground-truth estimates of the 1D depth map of these slices by performing the above epipolar-only imaging procedure for each of the 160 cosine-shifted patterns used in the previous experiments.

Figures 11-15 show reconstruction results, following exactly the same conventions as in Figures 6-10.

As a final epipolar-only imaging experiment, we repeated the above procedure for all epipolar lines on the image plane, using three six-pattern sequences: one derived by optimizing a six-pattern code matrix and one for six-pattern MPS and EPS patterns. Figure 16 shows the reconstruction results for a scene consisting of a bowl and the candle side by side.

¹In practice, we avoid discretization effects by projecting “thick” epipolar stripes that consist of 10 adjacent epipolar lines on the projector’s pixel grid.

²Even though we project 10-pixel wide epipolar stripes onto the scene, we do only extract one epipolar line from the captured image.

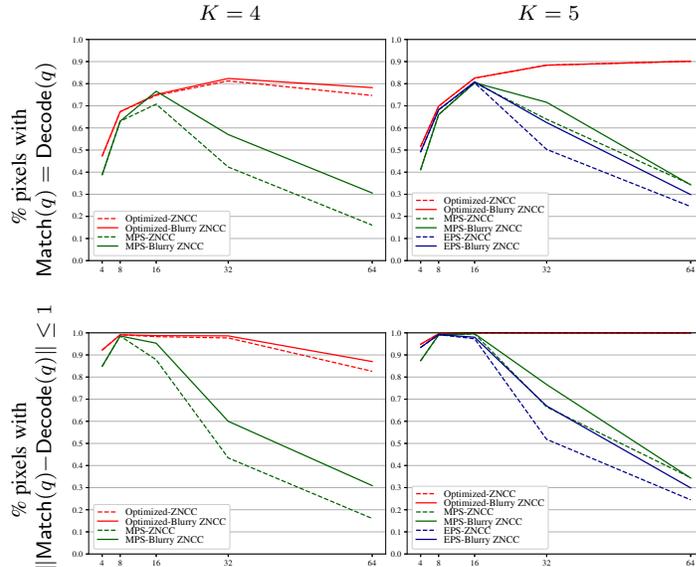


Figure 4: **ZNCC Decoding versus Defocused ZNCC Decoding.** Dashed lines correspond to standard ZNCC decoding where defocus blur is not taken into account. Solid lines show what happens when we use defocused ZNCC decoding instead. Each datapoint in these graphs represents a distinct 3D acquisition experiment.

D.3. Quantitative Evaluation of Defocused ZNCC Decoding

Our last experiment is a quantitative comparison of ZNCC decoding and defocused ZNCC decoding for the planar scene shown in Figure 7 (2nd row, 3rd column). All our codes were optimized without taking defocus into account, *i.e.*, we consider the effect of modeling defocus in the decoding stage only.

This experiment was conducted as follows. We positioned the planar scene at the midpoint of the working volume, manually focused the projector onto that plane and set the camera’s aperture to be small enough to ensure that the entire working volume was within the camera’s depth of field. We then displaced the planar scene to the far end of the volume, to the position shown in Figure 7 of the main paper. After projector calibration, the maximum size of its defocus blur kernel over the working volume was estimated to be approximately one pixel. Ground-truth depth was acquired using a 160-pattern sequence of shifted cosines, as explained in the previous experiments.³

We then optimized code matrices for $K = 4$ and $K = 5$ patterns, an error tolerance of zero pixel, and several frequencies. Figure 4 shows the reconstruction results, as a function of frequency, for both our optimized codes as well as MPS and EPS codes. Despite the small amount of projector defocus blur, our results show that defocused ZNCC decoding does have a significant impact on the results, especially for MPS and EPS codes, and especially for 4 patterns. The somewhat smaller performance improvement exhibited by our codes suggests that they are already quite robust to defocus, even without modeling this effect directly into the optimization.

References

- [1] M. Gupta and S. Nayar, “Micro Phase Shifting,” in *Proc. IEEE CVPR*, pp. 813–820, 2012.
- [2] D. Moreno, K. Son, and G. Taubin, “Embedded phase shifting: Robust phase shifting with embedded signals,” in *Proc. IEEE CVPR*, pp. 2301–2309, 2015.
- [3] P. Mirdehghan, W. Chen, and K. N. Kutulakos, “Optimal Structured Light a la Carte: Supplemental Document,” in *Proc. IEEE CVPR*, 2018. Also available at <http://www.dgp.toronto.edu/OptimalSL>.

³Since this sequence had cosines of very low frequency, we expect the ground-truth depth estimates to be robust to the small projector defocus exhibited in this experiment.

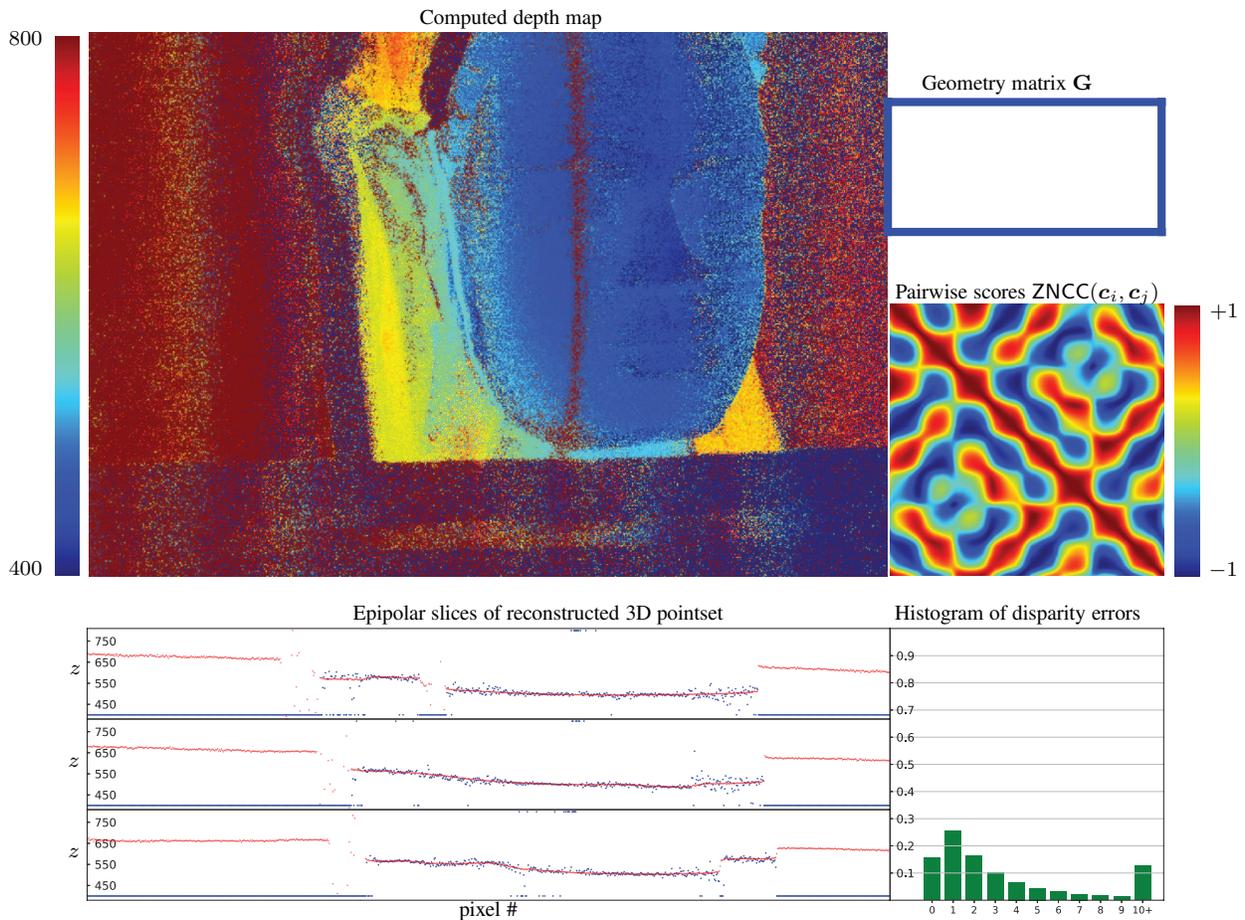


Figure 5: **Visualizing the results of a 3D acquisition experiment.** *Depth map:* Raw depth map computed in the experiment. Depth values are color-coded, with the darkest blue corresponding to a depth of 400 millimeters or less and darkest red corresponding to a depth of 800 millimeters or more. *Geometry matrix:* The matrix \mathbf{G} used for code optimization and ZNCC decoding in the experiment. In the example above the geometry matrix \mathbf{G} was all ones, *i.e.*, no geometry constraints were imposed during code optimization and during decoding. *ZNCC scores:* Pairwise scores of the columns of the code matrix used in the experiment. Dark red regions signify sets of code vectors that are very similar according to the ZNCC similarity score whereas dark blue regions represent sets of very dissimilar code vectors. *Epipolar slices:* We show a top view of three 2D slices of the reconstructed 3D pointset, along three different epipolar planes. For each slice, we plot the depth z of each pixel along the epipolar line, in millimeters. Blue points are the reconstructed 3D points on the slice; red points are the 3D points reconstructed by the “ground-truth” pattern sequence, which uses many more patterns. Points whose reconstructed depth is either smaller or larger than the indicated range of z values are plotted as 3D points at the minimum or maximum depth, respectively. We use this convention to improve legibility of the reconstructed surface, while also indicating the presence of outliers. Note that some of the “ground-truth” depths can be incorrect as well (*e.g.*, “floating” red points in the above plots that are far from the apparent cross-sections of the scene with the epipolar planes). *Histogram of disparity errors:* Here we quantify the difference between reconstructed and “ground-truth” disparities. Specifically, bin ϵ holds the fraction of pixels with known ground-truth disparity whose reconstructed disparity in this experiment differs by exactly ϵ pixels from their ground-truth value.

- [4] M. O’Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, “Homogeneous codes for energy-efficient illumination and imaging,” in *Proc. ACM SIGGRAPH Asia*, 2015.
- [5] M. O’Toole, J. Mather, and K. N. Kutulakos, “3D Shape and Indirect Appearance by Structured Light Transport,” *IEEE T-PAMI*, vol. 38, no. 7, pp. 1298–1312, 2016.

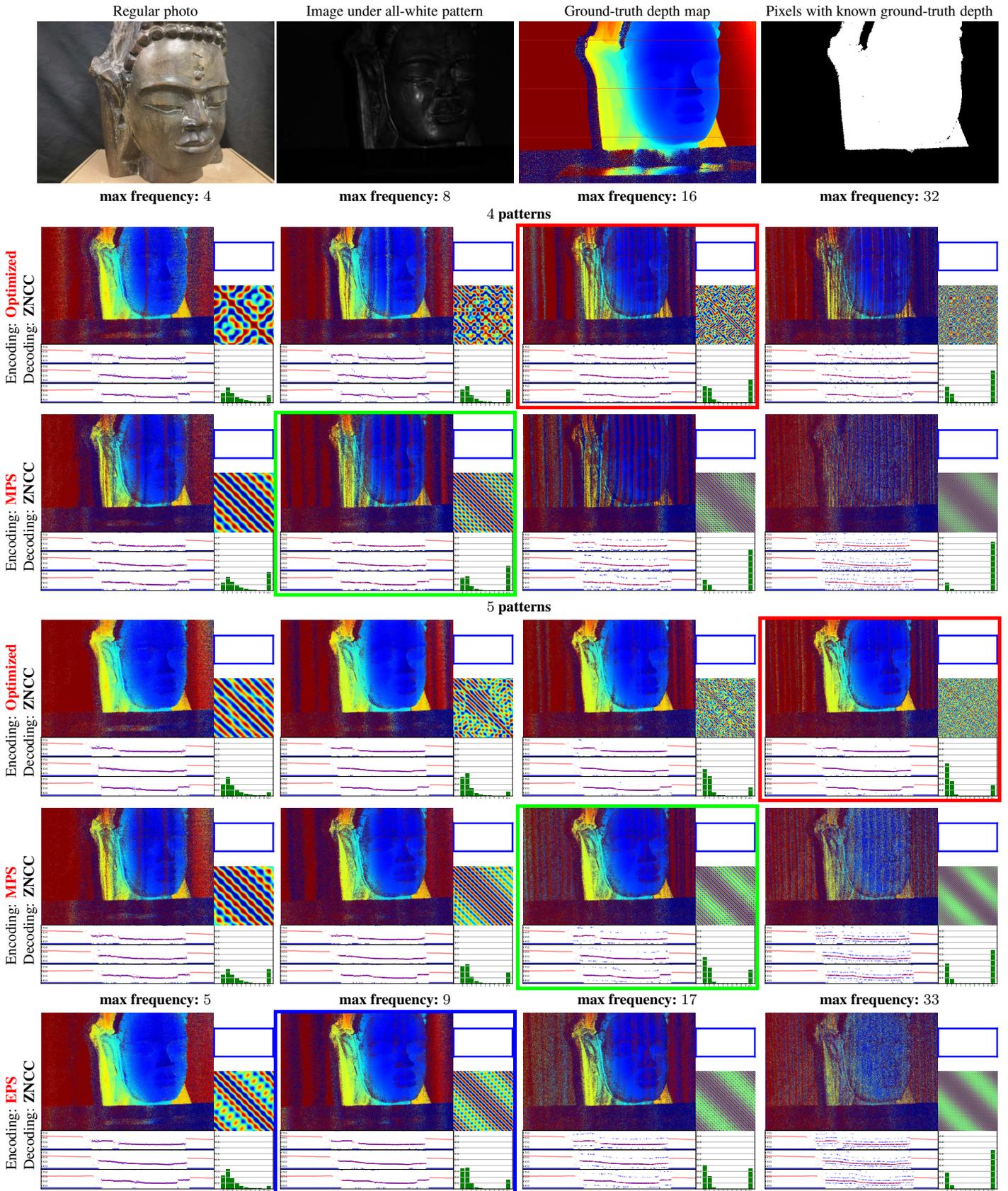


Figure 6: **Code performance as a function of maximum spatial frequency.** Please zoom into the electronic copy for details. In these experiments our codes were optimized for the exponential-plus-additive-Gaussian noise model; for exact disparity estimation (*i.e.*, error tolerance $\epsilon = 0$); and with no geometric constraints. To generate MPS codes, we set their maximum spatial frequency and use the MPS author-supplied code to generate the codes themselves. Embedded Phase Shifting (EPS) requires a minimum of five patterns. We use our own implementation of EPS code generation. For 5-pattern EPS sequences, their maximum spatial frequency is one unit higher than their power-of-two base frequency. Color frames indicate the best-performing codes for each case (red for ours, green for MPS, blue for EPS). The three epipolar lines corresponding to the epipolar slices shown are highlighted on the ground-truth depth map in the top row.

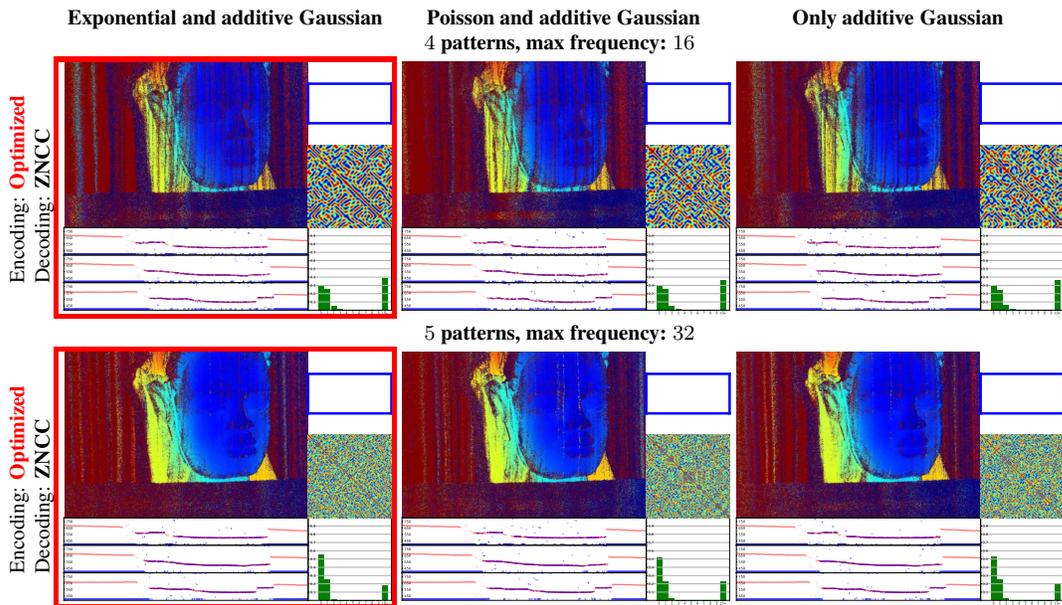


Figure 7: Code performance as a function of camera noise model. Please zoom into the electronic copy for details. Red-framed experiments are identical to the red-framed experiments in Figure 6; they are shown here for completeness.

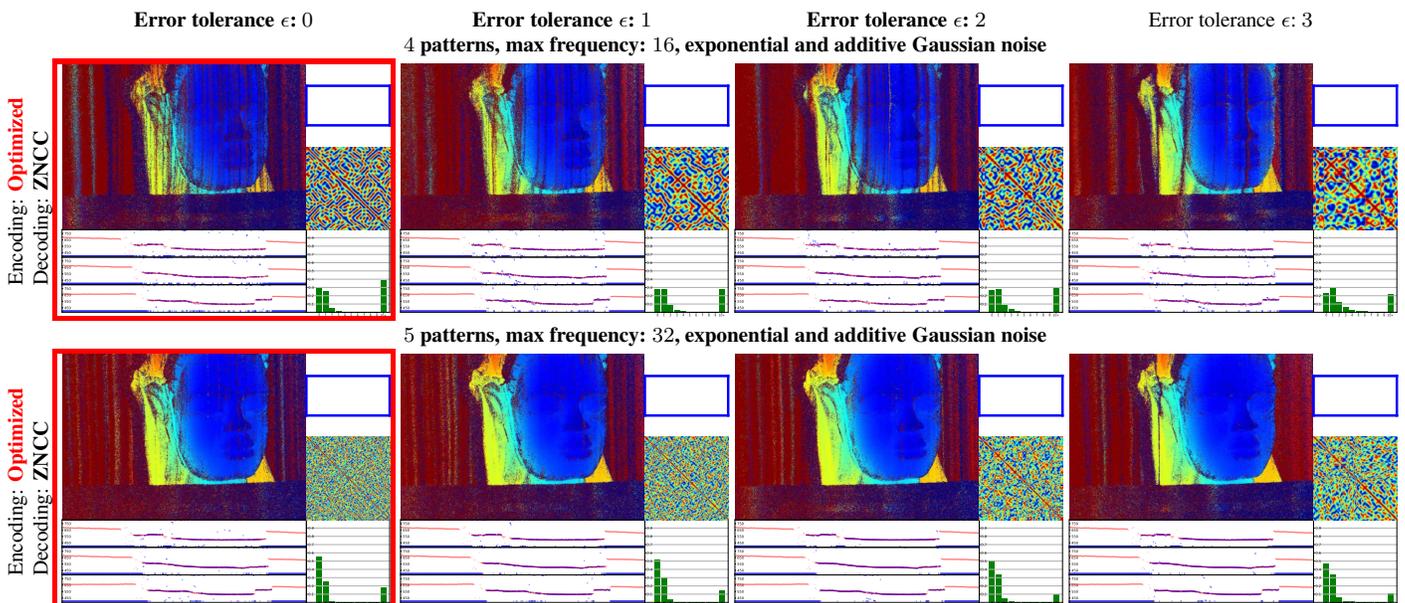


Figure 8: Code performance as a function of error tolerance. Please zoom into the electronic copy for details. Red-framed experiments are identical to the red-framed experiments in Figure 6; they are shown here for completeness.

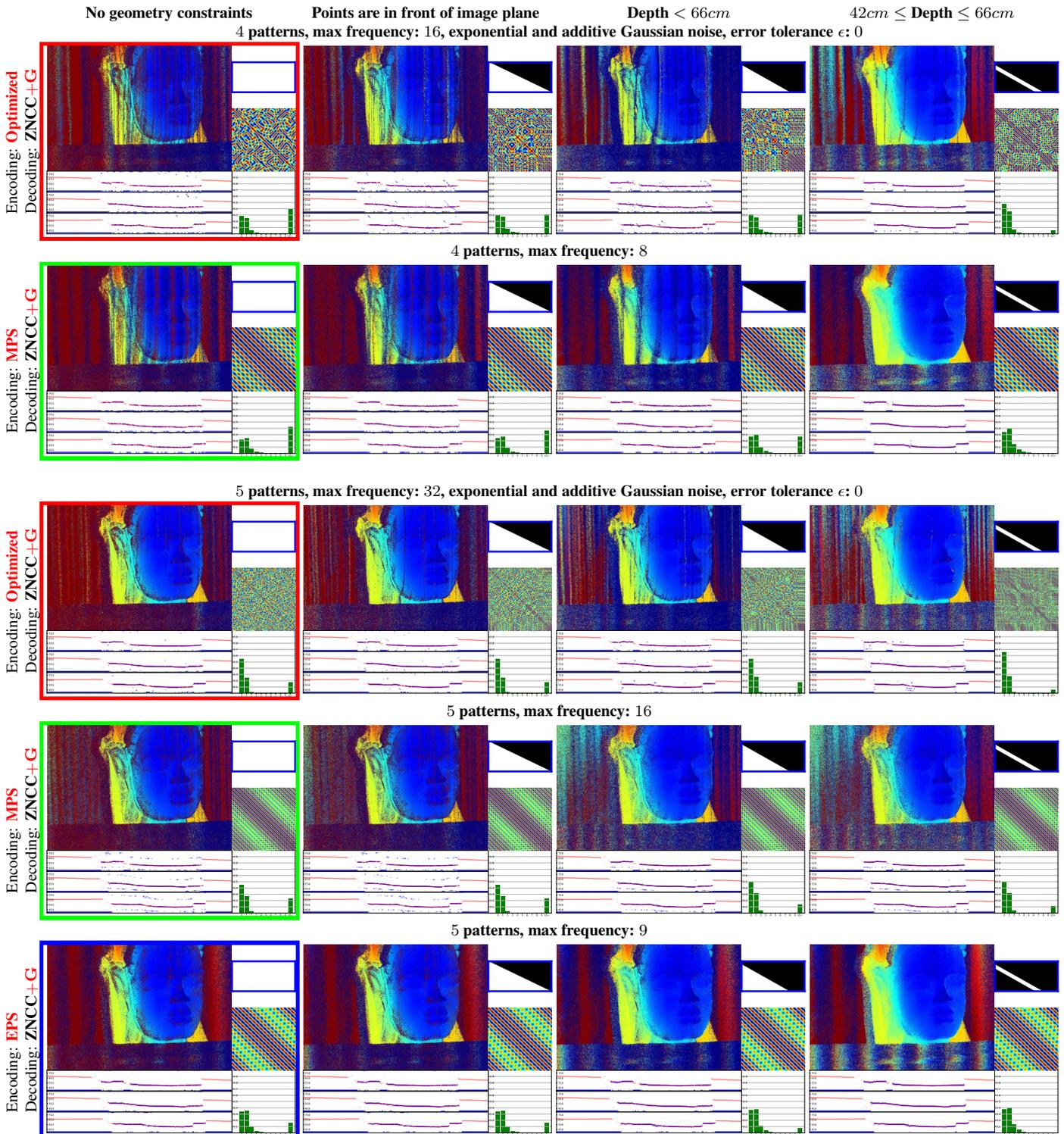


Figure 9: Code performance comparisons as a function of geometric constraints. Please zoom into the electronic copy for details. Red-, green- and blue-framed experiments are identical to the corresponding ones shown in Figure 6; they are shown here for completeness.

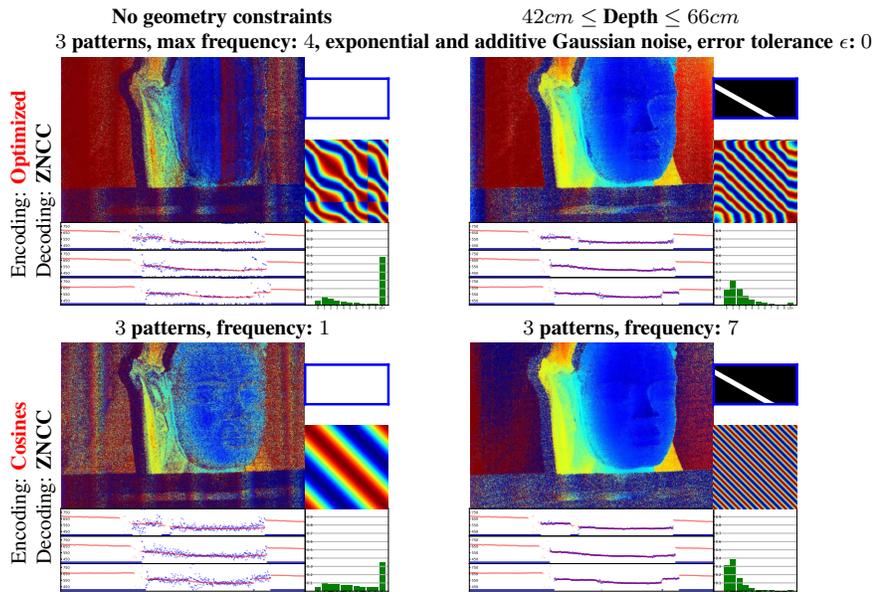


Figure 10: **Optimized 3-pattern code matrices versus 3-pattern phase shifting.** *Top row:* Reconstruction results for two optimized 3-pattern code matrices, both with and without geometry constraints. The results shown are for the best-performing maximum frequencies in the set $\{4, 8, 16, 32, 64\}$. *Bottom row:* For comparison, we also show reconstruction results for conventional 3-pattern phase shifting. In the absence of geometry constraints (left) a cosine of frequency of 1 is required to avoid phase-wrapping ambiguities. For the depth range we consider in the right column, ambiguities can be avoided with a cosine of maximum frequency 7.

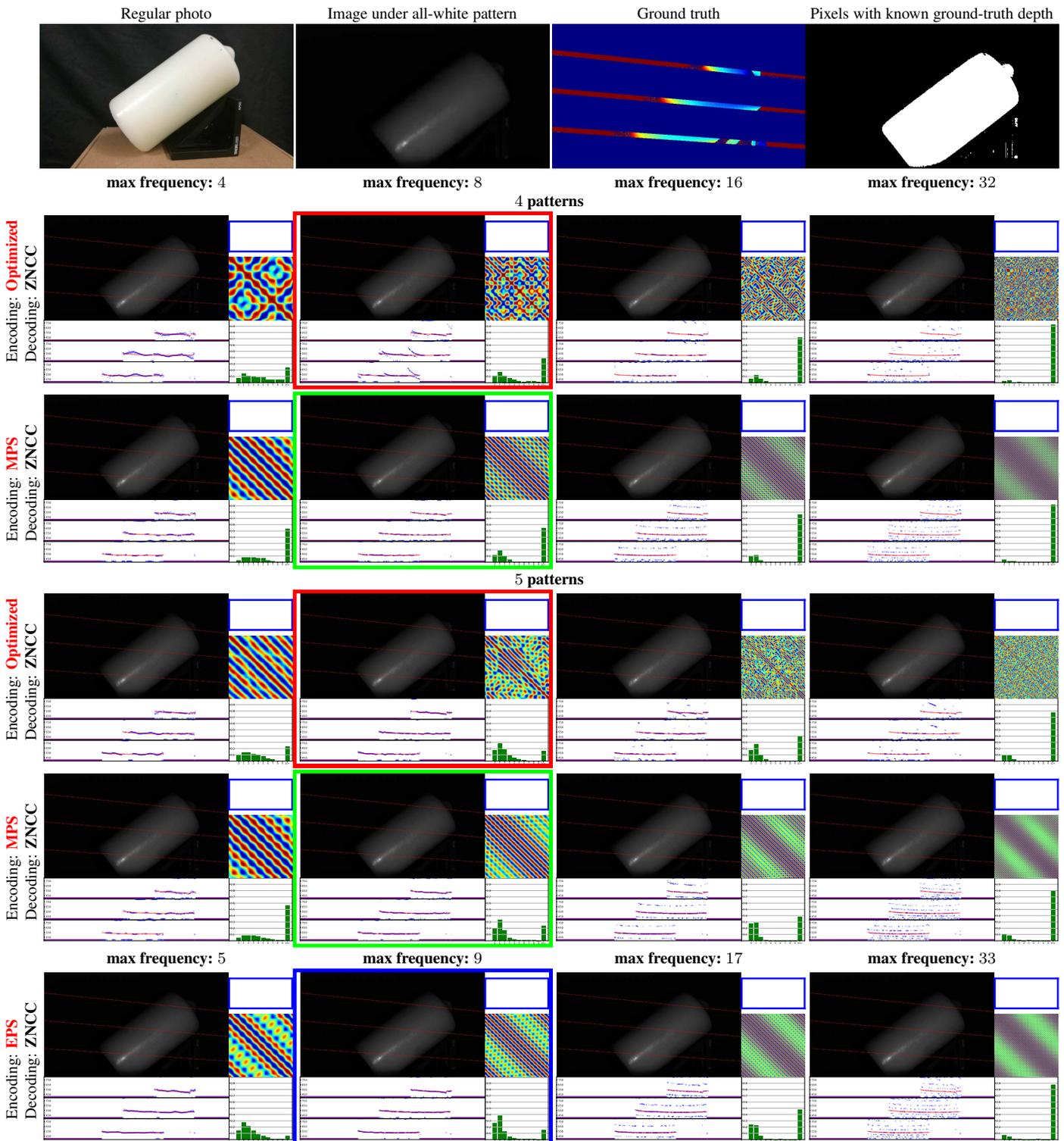


Figure 11: Code performance as a function of maximum spatial frequency. Since we only reconstructed three slices of the scene, a full depth map is not available. In lieu of that, we show an image of the scene with the three chosen epipolar lines superimposed.

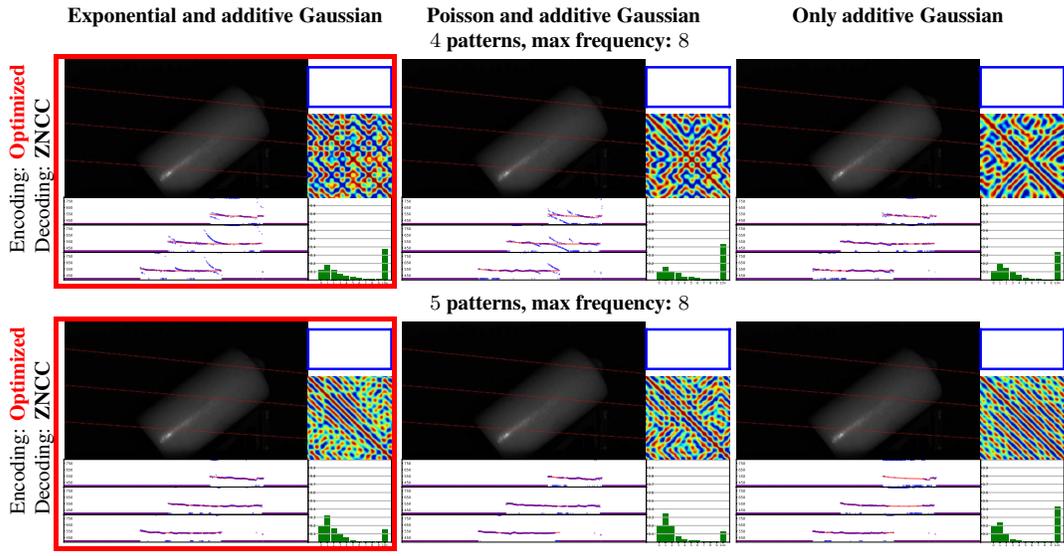


Figure 12: Code performance as a function of camera noise model.

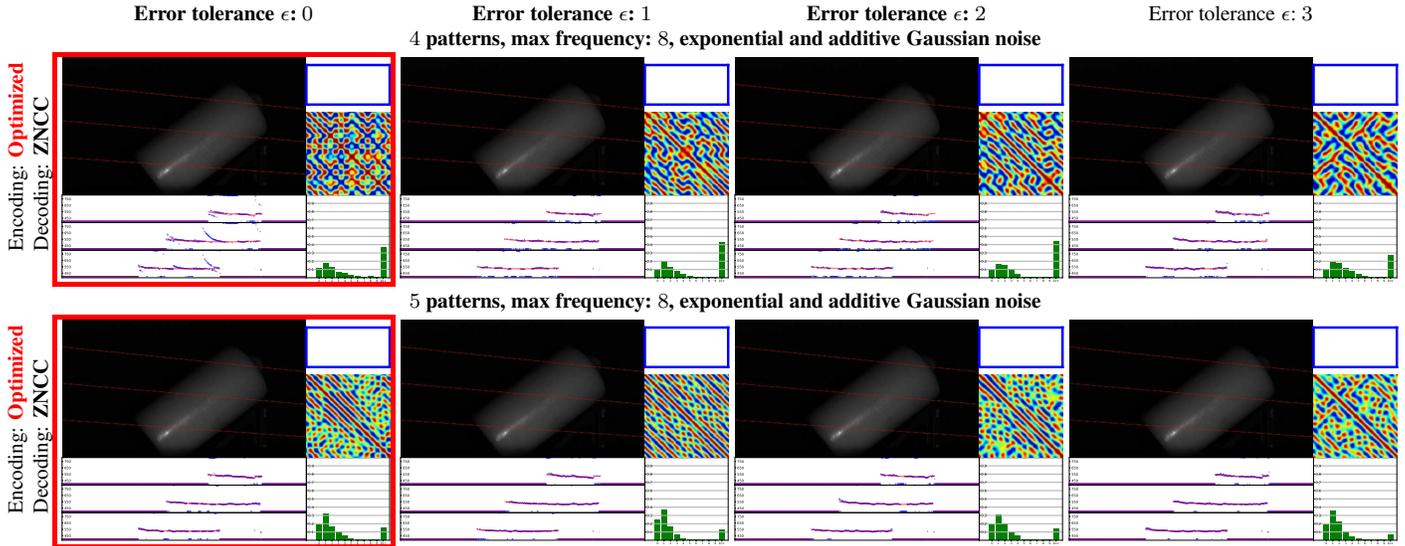


Figure 13: Code performance as a function of error tolerance.

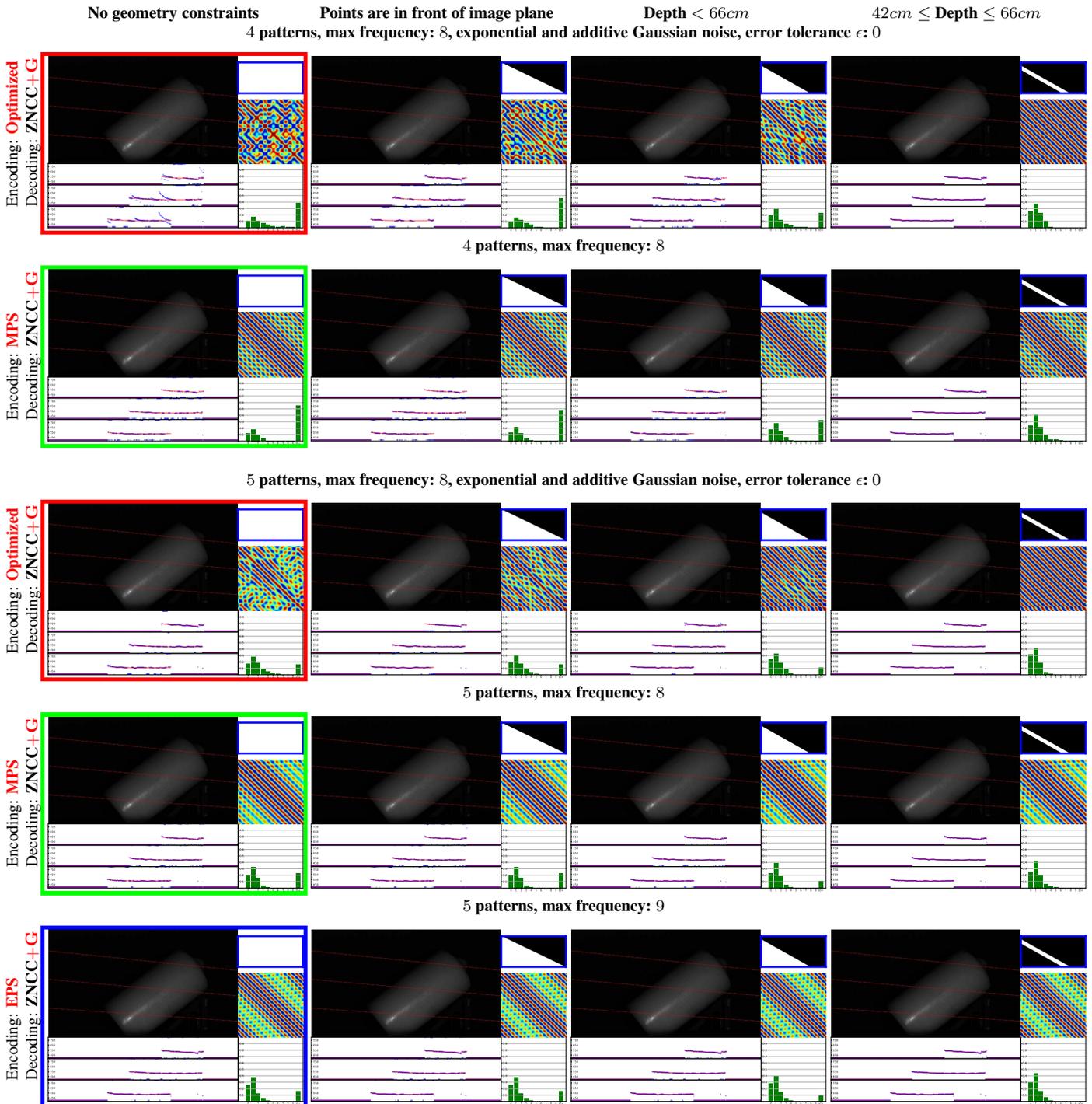


Figure 14: Code performance comparisons as a function of geometric constraints.

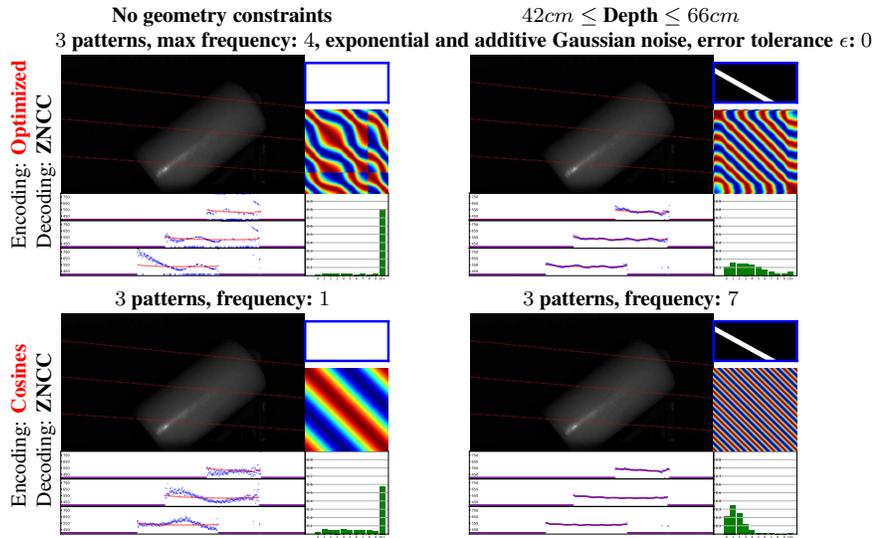


Figure 15: Optimized 3-pattern code matrices versus 3-pattern phase shifting.

Optimized 6 patterns, max frequency:16 MPS 6 patterns, max frequency:16 EPS 6 patterns, max frequency:16
 error tolerance $\epsilon: 1$

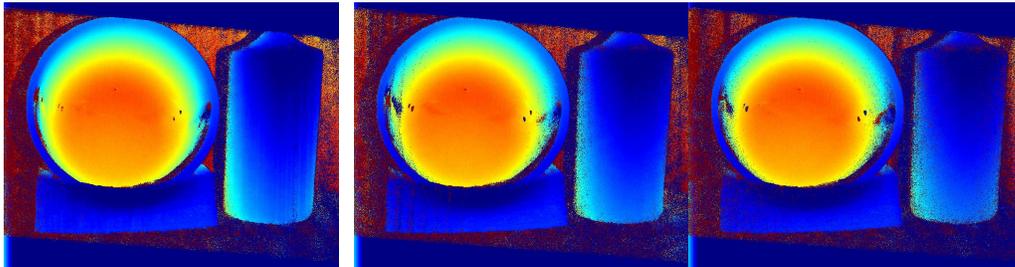


Figure 16: Reconstruction results for six-pattern sequences under epipolar-only imaging conditions. The candle is the same object shown in Figure 11. For a full set of experiments on the bowl refer to the project web page [3].